

Estudo sobre Séries Temporais e Machine Learning para a Identificação e Redução de Emissões de Gases de Efeito Estufa no Brasil

Elaine Thamires Pereira RA 10408284, Giovana Vieira da Silva RA 10408509

Faculdade de Computação e Informática (FCI)
Curso Tec. Ciência de Dados
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

Resumo. Este estudo analisa as emissões de gases de efeito estufa no Brasil, utilizando séries temporais como abordagem principal. O contexto das mudanças climáticas, que demandam ações mitigadoras eficazes, motivou o trabalho. O objetivo foi prever as emissões futuras e identificar padrões temporais para subsidiar a formulação de políticas públicas ambientais. A justificativa baseou-se na necessidade de compreender a evolução dessas emissões em diferentes setores da economia brasileira, permitindo uma gestão mais informada e alinhada aos compromissos climáticos globais.

A metodologia envolveu a utilização de dados históricos normalizados de emissões. Os modelos ARIMA, PROPHET E HOLT foram escolhidos como ferramentas de predição de séries temporais e os resultados obtidos destacam os setores de Mudança de Uso da Terra e Florestas. O modelo de Suavização de Holt foi o que obteve o melhor resultado, apresentando as melhores métricas de desempenho, com um MAE de 5.694, RMSE de 8.091 e MAPE de 0,2%. Ele capturou melhor os padrões e indicou um aumento lento nas emissões.

Palavras-chave: Análise de Séries Temporais; Aprendizado de Máquina; ARIMA; PROPHET; HOLT; Emissões de Gases de Efeito Estufa.

Abstract. This study analyzes greenhouse gas emissions in Brazil, using time series as the main approach. The context of climate change, which demands effective mitigation actions, motivated the work. The objective was to forecast future emissions and identify temporal patterns to support the formulation of environmental public policies. The justification was based on the need to understand the evolution of emissions across different sectors of the Brazilian economy, allowing for more informed management aligned with global climate commitments.

The methodology involved the use of normalized historical emission data. The ARIMA, PROPHET, and HOLT models were chosen as time series forecasting tools, and the results highlight the sectors of Land Use Change and Forests. The

Holt's Exponential Smoothing model produced the best results, achieving the highest performance metrics, with a MAE of 5.694, RMSE of 8.091, and MAPE of 0.2%. It captured the patterns more effectively, indicating a slow increase in emissions.

Keywords: *Time Series Analysis; ARIMA; PROPHET; HOLT; Machine Learning; Greenhouse Gas Emissions.*

1. Introdução

Os gases de efeito estufa são considerados pelo mundo um dos grandes agentes de mudanças climáticas globais, afetando ecossistemas, economias e comunidades ao redor do mundo. O Brasil, por sua vez, com economia emergente e com uma imensa área florestal, atua continuamente na tentativa de reduzir as emissões de gases de efeito estufa (GEE), apesar de ser o sexto maior emissor climático do mundo (Energia e Ambiente, 2023). Em 2021, o Brasil apresentou o segundo maior aumento de emissões de GEE em quase duas décadas, alcançando 2,4 bilhões de toneladas brutas. Dessa forma, esses dados demonstram a urgência de ações conjuntas para diminuir as emissões, especialmente nos setores mais críticos da economia brasileira (SEEG, 2023).

O aumento das emissões no Brasil está ligado a setores como agricultura, produção de energia e mudanças no uso do solo, que juntos constituem o centro das atividades econômicas que mais impactam as emissões de gases de efeito estufa no país. Essa situação também ameaça biomas nacionais, como a Amazônia e o Cerrado, que enfrentam pressões constantes e agravadas pelo crescimento das emissões (SEEG, 2023).

Entre as consequências das emissões de GEE estão os eventos climáticos extremos, como ondas de calor, secas intensas, inundações e furacões. Dados do Painel Intergovernamental sobre Mudanças Climáticas (IPCC, 2021) indicam que esses eventos podem desestabilizar sistemas alimentares, agravar a escassez de água, reduzir a biodiversidade e aumentar a incidência de doenças. Além disso, doenças transmitidas por vetores, como malária e dengue, podem se espalhar com maior rapidez devido ao aumento de temperatura, chuvas e umidade (Nahur, 2014).

Se o mundo quiser manter o aquecimento abaixo de 1,5°C em relação aos níveis pré-industriais, será necessário cortar pela metade as emissões de GEE até 2030. Com base nas Contribuições Nacionalmente Determinadas atuais, submetidas no âmbito do Acordo de Paris, as emissões de gases de efeito estufa levariam a um aquecimento de 2,7°C até o final do século, em um cenário que poderia ter implicações desastrosas para os ecossistemas e o bem-estar da sociedade. Segundo Radka (2023), as emissões de gases de efeito estufa causadas pelo homem ameaçam a saúde humana e ambiental, causando danos cada vez maiores e perigosos caso não seja feita uma forte intervenção climática.

A ciência climática estipula que os próximos anos marcarão um ponto de inflexão no futuro do planeta e qualquer habilidade em prever emissões setoriais futuras e identificar os principais contribuintes será primordial para orientar políticas eficientes. As ações de mitigação devem ser baseadas, portanto, em dados sólidos e análises detalhadas que permitam ao Brasil e ao restante do mundo alinhar suas rotas de desenvolvimento com os objetivos globais de sustentabilidade. Diante disso, é fundamental obter uma compreensão detalhada dos setores responsáveis pelas maiores emissões e antecipar

tendências futuras para guiar políticas públicas e nortear possíveis estratégias de redução mais eficazes.

Assim sendo, este projeto consiste em desenvolver ferramentas analíticas direcionadas à previsão de emissões futuras e à identificação dos setores que mais contribuem para essas emissões no Brasil. Com o uso de técnicas de aprendizado de máquina e análise de séries temporais, o objetivo é oferecer informações que auxiliem na elaboração de políticas públicas e estratégias voltadas para a redução das emissões.

As mudanças climáticas vêm causando transformações imprevisíveis e alcançando cada vez mais números alarmantes em todo o planeta. A urgência em mitigar os GEE vem sendo pauta de acordos internacionais e discussões éticas ao longo de anos. Segundo o relatório publicado em 2023 pela UNEP, o mundo vivencia uma preocupante aceleração em números de escala records de temperaturas. A previsão é que 2023 chegaria a temperaturas superiores a 1,5°C e chegou a alcançar 1,8°C na média térmica, sendo o mês de setembro de 2023 o mês mais quente registrado na história.

O aumento da temperatura global e as mudanças climáticas trazem uma série de problemas graves que afetam diretamente a sociedade e o meio ambiente (IPAM, 2024). Entre eles, destaca-se o aumento da frequência e intensidade de eventos climáticos extremos, como enchentes, tempestades e secas, impactando a segurança hídrica e alimentar, além de agravar desastres naturais. Outro problema significativo é a perda da cobertura de gelo nos polos, contribuindo para o aumento do nível do mar e alterações nos ecossistemas marinhos e terrestres. Mudanças na disponibilidade de recursos hídricos e na agricultura, como a diminuição das chuvas em regiões tropicais e subtropicais, também agravam a vulnerabilidade alimentar, principalmente em países em desenvolvimento.

Diante desse cenário, estudar meios de mitigar os danos causados pelas mudanças climáticas e pelas emissões de GEEs se torna uma obrigação de todos nós que vivemos em sociedade. A crescente pressão ambiental exige uma ação coletiva e consciente, que não apenas responsabilize governos e empresas, mas também cada cidadão em suas escolhas e hábitos diários. O avanço da ciência e da tecnologia nos proporciona ferramentas valiosas para compreender a dimensão desses problemas e nos motiva a desenvolver soluções cada vez mais eficazes. Portanto, é nosso dever, enquanto parte integrante da sociedade, buscar continuamente alternativas para reduzir nosso impacto ambiental e colaborar com a construção de um futuro mais sustentável e equilibrado.

Especificamente no contexto brasileiro, o país é o quinto colocado no ranking de países mais poluentes, e os dados mostram duas preocupantes tendências de aumento nas emissões de gases de efeito estufa. Desde 2010, quando a Política Nacional sobre Mudança do Clima estabeleceu metas de redução, o Brasil aumentou suas emissões em 23,2%, o que coloca o país em desvantagem para cumprir sua Contribuição Nacionalmente Determinada no Acordo de Paris. Apesar da recessão causada pela pandemia em 2020 onde atingiu uma redução significativa nas emissões de GEE, com um total de 1.675,76 milhões de toneladas de CO₂ equivalente (CO₂e), representando uma redução de 48,2% em relação às projeções iniciais de 3.236 milhões de toneladas para o ano, as emissões continuaram a subir, principalmente devido às atividades rurais, que representam 73% das emissões nacionais, e à especulação com terras, (Potenza et al. 2021).

Esse cenário destaca a necessidade de ações estratégicas e políticas públicas robustas para reduzir ainda mais as emissões nacionais. Justifica-se, portanto, a relevância deste projeto no contexto brasileiro, onde o estudo das mudanças climáticas e suas implicações regionais são fundamentais para o planejamento estratégico e a tomada de decisões por parte do poder público e da sociedade. A modelagem de séries temporais surge como uma ferramenta para fornecer informações sobre as tendências futuras de emissões de CO₂e, permitindo uma visão clara e científica sobre o comportamento das emissões ao longo do tempo.

Ao identificar setores econômicos que mais contribuem para as emissões de GEE, este estudo visa colaborar diretamente para desenvolvimento sustentável (ODS), alinhados às diretrizes da Organização das Nações Unidas (ONU) podendo fornecer subsídios para a criação de estratégias mais eficazes e direcionadas, auxiliando o Brasil a alcançar suas metas de redução de emissões e contribuindo para a preservação ambiental e o bem-estar da sociedade.

O objetivo deste projeto é utilizar técnicas de aprendizado de máquina para identificar os setores econômicos que mais contribuem para as emissões de gases de efeito estufa no Brasil para realizar uma previsão de seu potencial crescimento e, com base nas previsões obtidas, propor medidas de melhorias para políticas públicas, direcionando esforços e recursos para os setores mais críticos. Ao identificar os maiores emissores, as autoridades podem desenvolver estratégias mais eficazes para mitigar as emissões, investindo em novas tecnologias, promovendo um impacto significativo na luta contra as mudanças climáticas e contribuindo para o desenvolvimento sustentável do planeta.

Especificamente, realizar a Coleta, pré-processamento e exploração de dados no SEEG. Seleção de recursos a serem direcionados com modelos de aprendizado de máquina, que serão utilizados para a análise de emissões por setor. Construir um modelo de Machine Learning a partir de análises passadas para criar projeções da evolução das emissões nos principais setores produtivos. Avaliação e reavaliação do modelo final, incluindo ajuste de parâmetros e validação da precisão das previsões. Publicar e apresentar os resultados dos modelos em um formato compreensível para pesquisadores, tomadores de decisão e público. Discussão e propostas de melhorias e estratégias que poderiam ser consideradas para mitigar os impactos das emissões nos setores de maior contribuição.

2. Referencial Teórico

Para a Organização Mundial de Saúde (OMS), a exposição à poluição do ar é a maior responsável pela causa de milhões de mortes anuais, especialmente em países com menos recursos. De acordo com Silva e Arbilla (2022), atualmente, as doenças e a mortalidade causadas pela poluição do ar são consideradas o maior risco ambiental à saúde humana, comparáveis aos problemas decorrentes da alimentação inadequada e do tabagismo.

Os GEE são substâncias químicas que podem ser encontradas na atmosfera, que conseguem absorver calor da radiação infravermelha emitida pela superfície da Terra, por esse meio, contribuindo para o aquecimento do planeta (IPCC, 2021). De longe, os GEE abundantes incluem o dióxido de carbono (CO₂), o metano (CH₄), o óxido nitroso (N₂O) e os gases fluorados. Eles são medidos em termos de concentração em partes por milhão (ppm) ou partes por bilhão (ppb). A medição desses gases é indispensável para a

compreensão das mudanças climáticas, viabilizando a análise de séries temporais que mostram tendências e oscilações no tempo. A questão sobre a qualidade do ar está incluída nos ODS (Objetivos De Desenvolvimento Sustentável) principalmente no que diz respeito a saúde, energia, e qualidade de vida (SILVA E ARBILLA, 2022), logo, dados desse tipo são de grande relevância para a formulação de políticas ambientais, estratégias de mitigação relacionadas aos efeitos das mudanças climáticas e para estudos, como este, de análises envolvendo dados relacionados a temas de impactos ambientais.

No contexto da aplicação de séries temporárias para investigar a crescente preocupação com as mudanças climáticas e a necessidade de entender e prever as tendências de emissões de CO₂, o estudo de Yiqi Luo (2023) fornece uma análise comparativa interessante, onde o autor aplicou três modelos de previsão de séries temporais diferentes — ARIMA, Prophet e LSTM — para prever as emissões futuras de CO₂ com base em dados históricos de 1750 a 2017. Conforme indicado pelos menores erros absolutos médios (MAE), o modelo ARIMA apresentou um MAE de 1.86×10^{10} , o Prophet teve um MAE de 1.39×10^{10} enquanto o modelo LSTM foi o mais preciso, com um MAE de 1.26×10^{10} . Esses resultados indicam que, embora todos os modelos tenham mostrado previsões aceitáveis, o modelo LSTM, baseado em aprendizado de máquina, demonstrou maior precisão comparado aos modelos convencionais.

Outra abordagem interessante e também encontrada na literatura recente é a de Mattei e Cunha (2020), que investigam as emissões líquidas de GEE no Brasil no período de 1990 a 2017. Os dados utilizados foram coletados de diversas fontes, incluindo o Sistema de Estimativa da Emissão de Gases do Efeito Estufa (SEEG), o Instituto Brasileiro de Geografia e Estatística (IBGE) e o Departamento Nacional de Trânsito (DENATRAN). A metodologia aplicada inclui testes de raiz unitária, como Dickey-Fuller, Dickey-Fuller Aumentado, Phillips-Perron e ADF-GLS, para verificar a estacionariedade das variáveis. Após a identificação da ordem de integração, foram realizados testes de cointegração, como Johansen e Engle-Granger, que confirmaram a presença de cointegração entre as variáveis. Com as variáveis identificadas como cointegradas, foi estimado um modelo de correção de erros (VEC), permitindo capturar as relações de longo prazo e a dinâmica de curto prazo.

O trabalho de Bisognin e Werner (2018) também contribui para o embasamento deste estudo, ao explorar a dinâmica das emissões de CO₂ nos Estados Unidos, China e Brasil entre 1960 e 2013 e se baseia na coleta e análise de dados históricos de emissões de CO₂ per capita de três países extraídos da base de dados do World Bank Group. Utilizando tanto o modelo ARIMA quanto a suavização exponencial de Holt, o estudo dividiu os dados em períodos de ajuste e validação, garantindo uma avaliação mais rigorosa da acurácia dos modelos. Para a análise preliminar, os dados foram representados graficamente, facilitando a identificação de padrões e tendências nas séries temporais e acurácia dos modelos foi avaliada através de métricas como a média dos erros percentuais absolutos (MAPE) e a raiz quadrada da média dos erros quadráticos (RMSE). Os resultados esperados incluem previsões das emissões de CO₂ para os anos de 2020, 2025 e 2030, corroborando a importância da adoção de medidas efetivas na redução das emissões de gases de efeito estufa.

Concluimos que, com base nos estudos analisados, optamos por aplicar os modelos ARIMA, Prophet e Holt, à nossa base de dados para prever as emissões de gases de efeito estufa e identificar o setor com maior contribuição para essas emissões. A

escolha desses modelos se justifica pelos resultados promissores apresentados em estudos anteriores, que demonstraram sua eficácia na análise e previsão de séries temporais relacionadas às emissões de gases, possibilitando assim uma abordagem sólida e consistente para alcançar os objetivos desta pesquisa.

4. Metodologia

Segundo Morettin e Tolei (2004), o objetivo de uma série temporal é prever valores futuros da série, podendo ser em curto prazo, como em casos de séries de vendas, produção ou estoque, ou em longo prazo, como em séries de produtividade e outras variáveis de interesse. Além disso, uma análise de série temporal permite descrever o comportamento da série, identificar tendências, ciclos e variações sazonais, e é essencial para detectar periodicidades relevantes nos dados.

1. **Coleta e Organização dos Dados:** Os dados utilizados neste projeto foram obtidos pelo Sistema de Estimativas de Emissões de Gases de Efeito Estufa (SEEG) do Observatório do Clima. O SEEG considera diversos gases de efeito estufa como dióxido de carbono, metano, óxido nitroso e hidrofluorocarbonetos e apresenta seus parâmetros de emissão de gases por diferentes setores da economia. O conjunto de dados inicial continha inconsistências e valores ausentes que precisaram ser tratados no pré-processamento para garantir a integridade da análise. As emissões de CO₂ foram medidas em toneladas por setor, permitindo uma comparação clara entre os diferentes setores ao longo do tempo.
2. **Pré-processamento dos Dados:** O pré-processamento dos dados foi uma etapa crítica para garantir a qualidade e a adequação do conjunto de dados antes da aplicação dos modelos de previsão. Esta fase envolveu as seguintes sub-etapas:
3. **Limpeza de Dados:**
 - **Deteção e tratamento de valores ausentes:** foram identificados e removidos. Em alguns casos, a interpolação foi usada para preencher valores faltantes, especialmente quando as lacunas eram pequenas.
 - **Remoção de outliers:** Para evitar que valores extremos distorcessem as previsões, foi utilizada a técnica do desvio interquartil (IQR) para detectar e remover outliers. Outliers foram definidos como valores acima de $Q3 + 1.5 \times IQR$ ou abaixo de $Q1 - 1.5 \times IQR$.
4. **Normalização dos Dados:** Os dados foram normalizados para garantir que as emissões entre os diferentes setores fossem comparáveis, independentemente de escala. A normalização foi realizada utilizando o método Z-score, que transforma os dados para que tenham uma média zero e desvio padrão 1, permitindo que todos os setores fossem analisados sob a mesma escala.

4.1. Diagrama da Solução

O pipeline de dados foi organizado de maneira a fim de atingir os objetivos do trabalho atual e baseado nos artigos correlatos. Para isso, após a aquisição dos dados, procede-se com o pré-processamento, nele inclui a limpeza e tratamento de variáveis

faltantes, além da normalização da variável “emissao” e transformação de variáveis (dados categóricos), se necessário.

Em seguida, será realizada a análise exploratória desses dados, com o intuito de identificar tendências e o comportamento das features, através da aplicação de estatísticas básicas. Nesta etapa também é incluída a plotagem de alguns gráficos para melhor compreensão e relação dos dados.

Depois, para a modelagem, as técnicas abrangem modelos estatísticos e modelos de aprendizado de máquina, sendo elas: suavização Exponencial de Holt, ARIMA e Prophet. Nesse sentido, com base nos trabalhos referenciados, o propósito é comparar o desempenho desses modelos, tendo em vista o tipo de série temporal retratada, com características de crescimento exponencial, sem estacionalidade e sem sazonalidade.

Por último, a avaliação dos modelos será realizada utilizando métricas como RMSE, MAE e MAPE, que permitirão medir o desempenho e a precisão das previsões geradas. Em resumo, as conclusões finais envolverão discussões sobre os resultados obtidos, além de sugestões para trabalhos futuros, culminando em uma síntese das descobertas do projeto.

4.2. Pipeline da Solução

O pipeline de dados foi organizado de maneira a fim de atingir os objetivos do trabalho atual e baseado nos artigos correlatos. Para isso, após a aquisição dos dados, procede-se com o pré-processamento, nele inclui a limpeza e tratamento de variáveis faltantes, além da normalização da variável “emissao” e transformação de variáveis (dados categóricos), se necessário.

Em seguida, será realizada a análise exploratória desses dados, com o intuito de identificar tendências e o comportamento das features, através da aplicação de estatísticas básicas. Nesta etapa também é incluída a plotagem de alguns gráficos para melhor compreensão e relação dos dados.

Depois, para a modelagem, as técnicas abrangem modelos estatísticos e modelos de aprendizado de máquina, sendo elas: suavização Exponencial de Holt, ARIMA e Prophet. Nesse sentido, com base nos trabalhos referenciados, o propósito é comparar o desempenho desses modelos, tendo em vista o tipo de série temporal retratada, com características de crescimento exponencial, sem estacionalidade e sem sazonalidade.

Por último, a avaliação dos modelos será realizada utilizando métricas como RMSE, MAE e MAPE, que permitirão medir o desempenho e a precisão das previsões geradas. Em resumo, as conclusões finais envolverão discussões sobre os resultados obtidos, além de sugestões para trabalhos futuros, culminando em uma síntese das descobertas do projeto.

4.3. Descrição da Base de Dados

A base de dados foi obtida através do site Base dos Dados, uma organização sem fins lucrativos e open source. Os dados explorados pertencem ao Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa, o SEEG, o qual é uma aplicação do

Observatório do Clima (OC) que apresenta estimativas anuais detalhadas das emissões nos principais setores responsáveis pela poluição climática no país.

O conjunto de dados dispõe de 454.850 registros no total e uma cobertura temporal datada entre 1970 e 2019, abrangendo 12 variáveis listadas abaixo:

“ano” — Representa o ano em que os dados foram coletados. “nivel_1” a “nivel_6” — Essas variáveis classificam as emissões em diferentes níveis setoriais. Começando pelo setor mais amplo até uma descrição mais detalhada em cada setor, facilitando a análise hierárquica das emissões. Exemplo: “nivel_1” sendo energia, agropecuária, mudança de uso da terra e floresta, já “nivel_6” sendo arroz, gado de corte, caminhões, embarcações, etc. “tipo_emissao” — Refere-se ao tipo de emissão de GEE. “gas” — Refere-se ao tipo de gás de efeito estufa emitido, como dióxido de carbono (CO₂), metano (CH₄), óxido nitroso (N₂O), etc. “atividade_economica” — Relaciona-se às atividades econômicas que contribuem para as emissões, como agricultura, transporte ou indústria, por exemplo, produção de combustível, cimento, pecuária, transporte de passageiros e de carga, entre outras. “produto” — Identifica o produto resultante das atividades econômicas relacionadas à emissão, como carne, alimentos e bebidas, aço, entre outros. “emissao” — Variável numérica que quantifica, em toneladas, o volume total de GEE que entra na atmosfera. Em relação à confiabilidade e à precisão desses dados, é válido enunciar que a metodologia do SEEG segue as recomendações do Inventário Brasileiro de Emissões e Remoções Antrópicas de Gases de Efeito Estufa, elaborado pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), e a partir das orientações baseadas no Painel Intergovernamental sobre Mudanças Climáticas (IPCC). A construção das estimativas de emissões é baseada em dados obtidos de diversas fontes: relatórios de entidades governamentais, instituições de pesquisa, agências setoriais e organizações não governamentais. Quando disponíveis, o SEEG utiliza fontes de dados públicas e gratuitas para a transparência do processo de cálculo e permite a qualquer interessado a reprodução dos mesmos. Contudo, essa abordagem metodológica confere à estimativa padrões internacionais e adequação às especificidades brasileiras.

4.4. Modelo ARIMA

Foi considerado o estudo do modelo Auto Regressive Integrated Moving Average (ARIMA) devido à sua robustez na modelagem de séries temporais, sua ampla divulgação e utilização (MORETTIN e TOLOI, 2004) e capacidade de capturar tanto padrões passados quanto variações aleatórias na série. Esse modelo é composto de três componentes: o autorregressivo (AR), que utiliza uma relação linear entre valores passados e o valor atual da série; a integração (I), que diferencia a série temporal para torná-la estacionária, eliminando tendências de longo prazo; e a média móvel (MA), que ajusta o valor atual com base nos erros de previsões passadas. A fórmula geral do modelo ARIMA é expressa como:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Onde:

- Y_t é o valor previsto da série temporal no instante t ;
- ϕ_i são os coeficientes de autoregressão;
- θ_i são os coeficientes da média móvel;

- e_t é o erro no tempo t ;
- ϵ_t é um termo de erro aleatório.

Para cada setor, foi testada uma combinação de parâmetros ARIMA (p, d, q), onde:

- p : Ordem do componente autoregressivo;
- d : Número de diferenciações aplicadas para tornar a série estacionária;
- q : Ordem do componente de média móvel.

Para aplicar o ARIMA corretamente, é essencial que a série seja estacionária, o que implica que suas propriedades estatísticas, como média e variância, devem ser constantes ao longo do tempo. A verificação da estacionariedade foi realizada por meio do Teste de Dickey-Fuller Aumentado (ADF), que testa a hipótese nula de que a série possui uma raiz unitária (não é estacionária). Quando o p -valor obtido é menor que 0,05, a hipótese nula é rejeitada, indicando que a série é estacionária e, portanto, adequada para o uso do ARIMA. Caso a série não fosse estacionária, aplicou-se a diferenciação de primeira ordem, que consiste em calcular a diferença entre observações consecutivas, removendo tendências de longo prazo e, assim, tornando a série estacionária. No caso deste projeto, o valor do parâmetro d foi fixado em 1 para diferenciar a série uma vez. A ordem dos outros parâmetros foi escolhida com base na análise gráfica das funções de autocorrelação parcial (PACF) e autocorrelação (ACF) da série temporal, e foram testadas várias combinações para definir a melhor configuração do modelo. A combinação de parâmetros ARIMA selecionada foi a (1,1,1), onde $p=1$ representa a ordem do componente autoregressivo, $d=1$ representa a diferenciação de primeira ordem aplicada para a estacionariedade, e $q=1$ representa a ordem do componente de média móvel. A escolha final foi determinada pelo Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC), critérios que penalizam a complexidade do modelo e auxiliam na escolha de uma configuração que ofereça o melhor ajuste aos dados com o mínimo de complexidade. O modelo ARIMA(1,1,1) apresentou baixos valores de AIC e BIC, e foi selecionado para a previsão de emissões de CO₂ para cada setor analisado.

No início da modelagem com o ARIMA, foram realizadas diversas etapas de pré-processamento para garantir que os dados estivessem adequados ao modelo. Primeiramente, o conjunto de dados foi carregado e analisado, com destaque para a identificação das variáveis relevantes, como os anos (dimensão temporal) e os valores de emissões de gases (variável a ser prevista). Em seguida, os dados foram convertidos para os tipos apropriados: a coluna de anos foi ajustada para o formato `datetime`, enquanto os valores de emissões foram convertidos para o tipo numérico, assegurando a compatibilidade com os cálculos exigidos pelo modelo.

Em seguida, foi realizado o teste ADF para verificar a estacionariedade da série temporal de emissões normalizadas. Esse teste foi aplicado diretamente à coluna `emissao_normalizada`, que continha os valores ajustados e normalizados de emissões de gases de efeito estufa para os anos analisados. O teste foi aplicado à série temporal utilizando a função `adfuller` da biblioteca `statsmodels` e os Resultados Obtidos foram: Estatística ADF: -30.1665, p -value: 0.0. Com base no valor- p abaixo de 0,05, a hipótese nula foi rejeitada, confirmando que a série é estacionária.

Figura 1 - Teste ADF

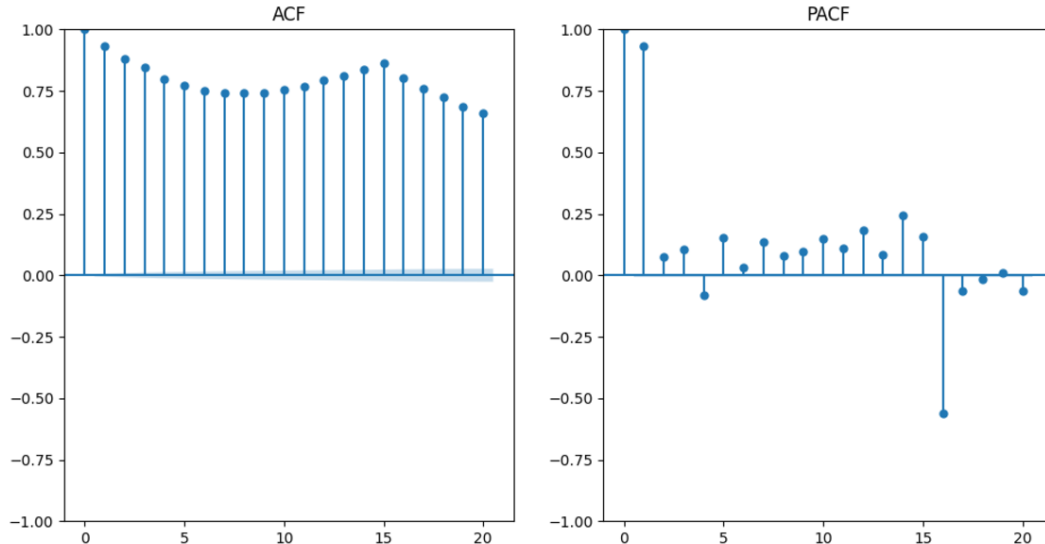
```
# Teste ADF
result = adfuller(df_modelo['emissao_normalizada'])
print('Estatística ADF:', result[0])
print('p-value:', result[1])

Estatística ADF: -30.166499056634994
p-value: 0.0
```

Fonte: Os autores (2024)

Para configurar o modelo ARIMA, foram realizados testes para definir os valores mais apropriados dos parâmetros p , d e q . O processo envolveu a aplicação de uma análise de autocorrelação e autocorrelação parcial para identificar a quantidade ideal de lags a serem utilizados. O gráfico ACF (Figura 2) mostrou como os valores da série estavam correlacionados com os lags anteriores e o gráfico (Figura 2) ajudou a isolar a influência de cada lag sobre os valores futuros, eliminando os efeitos intermediários. Ambos os gráficos indicaram padrões de autocorrelação significativos, permitindo determinar p e q iniciais de (3,0,3), que posteriormente foi confirmado como a melhor combinação.

Figura 2 - Gráficos ACF e PACF - Autocorrelação



Fonte: Os Autores (2024)

Os loops aninhados testam todas as combinações de p e q , ajustando um modelo ARIMA para cada par de valores. Durante cada iteração, o modelo é ajustado à série temporal e seu AIC é calculado. O AIC é uma medida que indica a qualidade do modelo, onde valores menores são preferíveis. Se o AIC do modelo atual for menor do que o melhor AIC previamente registrado, as variáveis que armazenam o melhor modelo e seus parâmetros são atualizadas.

Ao final da execução, o código imprime a melhor combinação de parâmetros encontrada, bem como o AIC correspondente. Os resultados obtidos ao ajustar o modelo

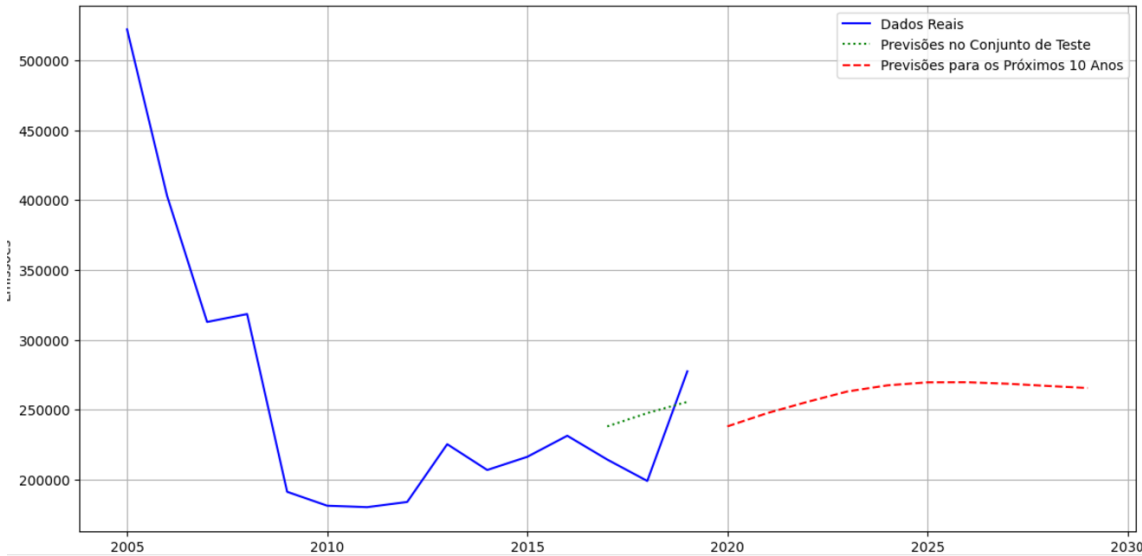
ARIMA indicam que a melhor combinação de parâmetros encontrada foi ARIMA(3, 0, 3), com um AIC de 1858967.37. Este valor de AIC sugere que o modelo se ajusta bem aos dados de emissões normalizadas, minimizando a perda de informação ao mesmo tempo em que penaliza a complexidade do modelo.

O resumo do modelo revela informações cruciais sobre os coeficientes do modelo, indicando a relação entre as observações passadas e a série temporal. Os coeficientes dos termos autoregressivos (ar.L1, ar.L2, ar.L3) são todos significativos, com os valores de -0.6568, 0.5531 e 0.9103, respectivamente. Esses coeficientes mostram como as emissões em momentos anteriores influenciam as emissões atuais. Já os coeficientes dos termos de média móvel (ma.L1, ma.L2, ma.L3) também são significativos, sugerindo que os erros de previsão em períodos passados impactam as previsões atuais.

A estatística Ljung-Box e o teste de Jarque-Bera fornecem insights adicionais sobre a adequação do modelo. O teste de Ljung-Box indica que não há autocorrelação significativa nos resíduos do modelo, enquanto o teste de Jarque-Bera sugere que os resíduos não seguem uma distribuição normal, o que pode requerer ajustes futuros.

Ajustando o modelo para 80% dos dados de treinamento e 20% de teste. Após definir os parâmetros, o modelo ARIMA foi treinado com os dados históricos para cada setor de emissões. Esta etapa foi realizada para ajustar o modelo aos dados, ou seja, aprender os padrões e comportamentos das emissões de GEE em cada setor. Durante o treinamento, o modelo utilizou as observações passadas para calibrar seus coeficientes e gerar previsões que seguem a mesma estrutura temporal observada.

Figura 3 - Previsões com Arima



Fonte: Os Autores (2024)

Tabela 1 - Resultados das Previsões com Arima

Ano: 2020	Previsão de Emissão: 237977.11
Ano: 2021	Previsão de Emissão: 247503.26

Ano: 2022	Previsão de Emissão: 255540.23
Ano: 2023	Previsão de Emissão: 262975.75
Ano: 2024	Previsão de Emissão: 267356.16
Ano: 2025	Previsão de Emissão: 269509.44
Ano: 2026	Previsão de Emissão: 269568.51
Ano: 2027	Previsão de Emissão: 268514.46
Ano: 2028	Previsão de Emissão: 266945.94
Ano: 2029	Previsão de Emissão: 265454.36

Fonte: Os Autores (2024)

Primeiro, a estacionariedade foi comprovada pelo teste ADF:

Estatística ADF: -30,166

p-value: 0.0.

Assim, esses resultados indicam que esta série é, de fato, estacionária. Assim, alguns parâmetros de p e q foram testados para ajustar o modelo. De qualquer forma, as métricas de desempenho obtidas foram piores do que o esperado:

RMSE: 592,94

MAE: 150,99

R²: -0,04

Uma possível explicação para esses resultados é que o modelo ARIMA não pode se ajustar adequadamente aos dados devido a flutuações não bem capturadas pelos componentes autorregressivos e de média móvel, ou simplesmente devido à complexidade interna na natureza dessas emissões, que podem ser não lineares e, portanto, não previsíveis. Apesar disso, o modelo foi utilizado para previsões para os próximos 10 anos, de 2020 a 2029, apresentando um cenário de leve aumento na série até o ano de 2026, seguido de queda gradual até 2029.

Os valores previstos (em toneladas) obtidos são:

2020: 237977.11

2025: 269509.44

2029: 265454.36

Esses valores sugerem uma primeira tendência de crescimento, talvez relacionada ao resumo das atividades econômicas ou à variação da política ambiental, depois estabilização e leve queda nos anos seguintes, observável no gráfico acima. O modelo é limitado do ponto de vista de desempenho, mas dá uma boa visão geral para uma avaliação preliminar das tendências futuras das emissões.

4.5. Prophet

O Prophet é uma biblioteca desenvolvida pelo Facebook que facilita a criação de modelos para previsão de séries temporais. Seus principais pontos fortes são a simplicidade de uso e a documentação clara. Trata-se de um modelo aditivo que também lida com sazonalidades e tendências de longo prazo. O modelo subjacente do Prophet é aditivo, ou seja, a previsão é construída pela soma de três componentes principais: tendência, sazonalidade e efeitos de eventos, com a inclusão de um termo de erro para capturar variações não explicadas pelos outros componentes.

A tendência é modelada de forma linear ou logística, dependendo do comportamento dos dados ao longo do tempo. A tendência linear é expressa pela seguinte equação:

$$g(t) = \alpha + \beta t$$

Onde α é a interceptação e β é a taxa de crescimento. No caso de dados que apresentam um comportamento de saturação, como uma tendência que se aproxima de um limite superior ou inferior, utiliza-se um modelo logístico, dado por:

$$g(t) = \frac{C}{1 + \exp(-(k(t - t_0)))}$$

Onde C é o valor máximo (limite superior), k é a taxa de crescimento e t_0 é o ponto de inflexão. Esse modelo permite que o Prophet capture comportamentos não lineares que são comuns em dados ambientais, como as emissões GEEs que podem se aproximar de limites de emissão devido a regulamentos ou inovações tecnológicas (Taylor & Letham, 2018).

A sazonalidade é modelada por meio de uma série de funções trigonométricas, representadas pela seguinte fórmula:

$$s(t) = \sum_{i=1}^N (a_i \cos(2\pi f_i t + \phi_i))$$

Onde f_i são as frequências das sazonalidades (por exemplo, sazonalidade anual, semanal, etc.), a_i são as amplitudes associadas a cada sazonalidade, e ϕ_i são os deslocamentos das fases sazonais. Esse componente permite capturar padrões cíclicos nos dados, como variações sazonais nas emissões de GEE causadas por fatores como a demanda por energia ou alterações nas atividades econômicas (Harvey & Peters, 1990).

O modelo do Prophet também se destaca pela sua flexibilidade. Ele permite ajustes intuitivos nos parâmetros, tornando-o acessível para analistas com pouco conhecimento em estatísticas ou modelagem de séries temporais. Isso o torna uma ferramenta valiosa para a previsão de emissões de gases de efeito estufa, pois possibilita

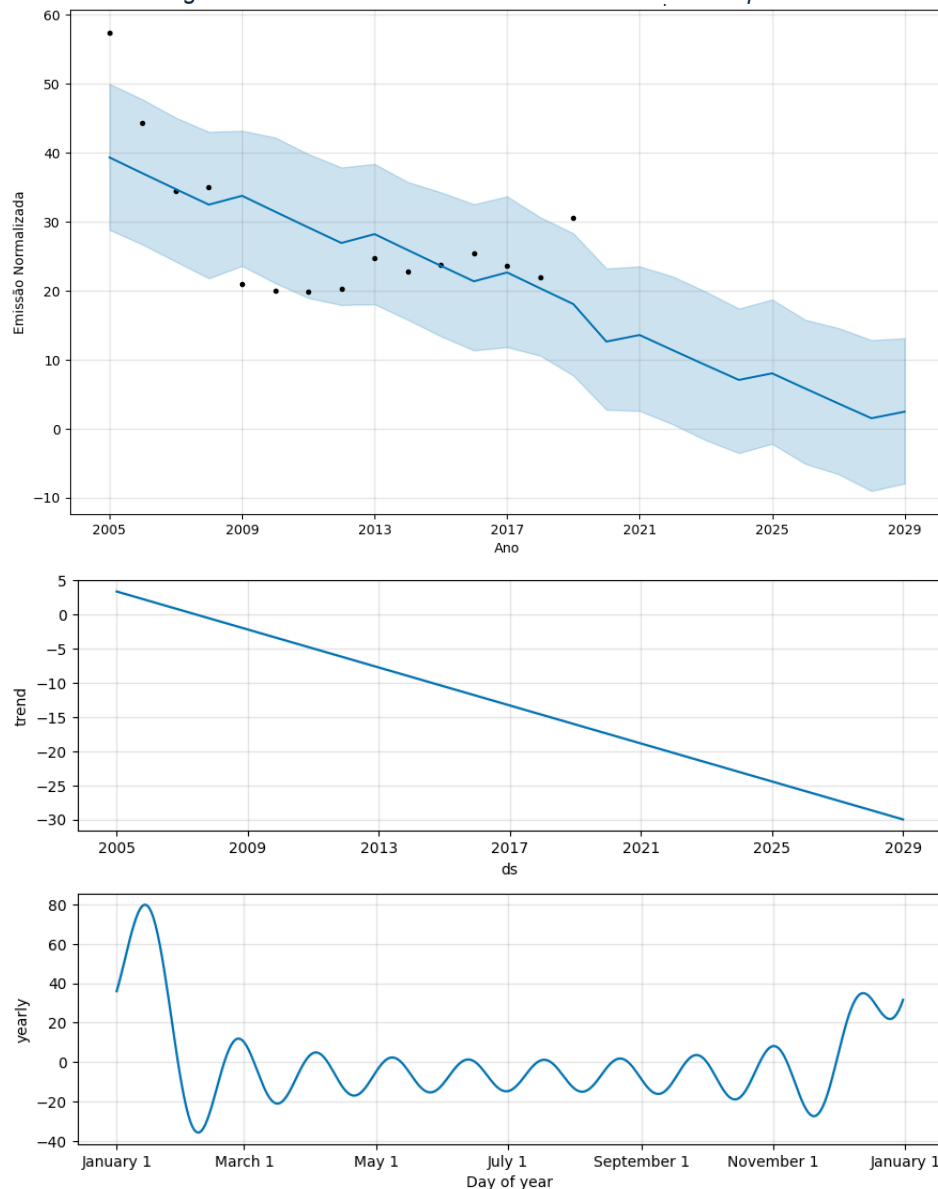
modelar dados de diferentes setores e ajustar os parâmetros de forma personalizada, conforme as características específicas de cada conjunto de dados (Taylor & Letham, 2018).

O processo com o Prophet iniciou-se com a preparação dos dados, onde as emissões normalizadas foram agrupadas por ano, calculando-se a média anual para cada ano presente no conjunto de dados. O DataFrame resultante foi ajustado para atender às exigências do modelo, renomeando a coluna "ano" para "ds" e "emissao_normalizada" para "y". Após essa etapa, o modelo foi treinado com os dados preparados, gerando previsões para os próximos 10 anos.

Os dados foram divididos em conjuntos de treinamento e teste, com 80% dos dados alocados para o treinamento do modelo e os 20% restantes utilizados para validação. O modelo foi treinado com os dados de treino e, em seguida, fez-se a previsão para o período correspondente aos dados de teste. Para avaliar a precisão do modelo, foram calculadas três métricas de desempenho: o erro absoluto médio (MAE), a raiz do erro quadrático médio (RMSE) e o erro percentual absoluto médio (MAPE). Essas métricas permitiram uma avaliação quantitativa do desempenho do modelo em relação aos dados reais.

Por fim, a análise das previsões foi complementada com gráficos dos componentes do modelo, incluindo as variações de tendência e sazonalidade, além de uma análise quantitativa das métricas de desempenho:

Figura 4 - Previsões de Emissão Normalizada com Prophet



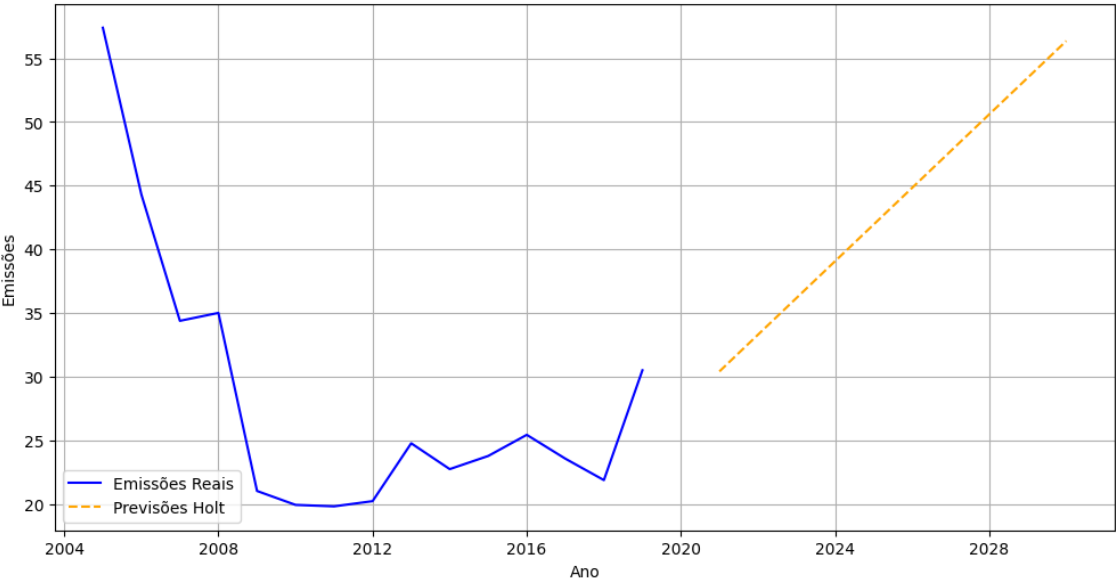
Fonte: Os Autores (2024).

4.6. Modelo Holt

O modelo de suavização exponencial de Holt, também conhecido como método de Holt-Winters, é uma extensão da suavização exponencial simples, projetada para lidar com séries temporais que apresentam tendência. Ao contrário do modelo simples, que é adequado para séries temporais sem tendência ou sazonalidade, o modelo de Holt permite modelar séries com tendência linear, ajustando a previsão com base no comportamento observado ao longo do tempo. O modelo utiliza dois componentes principais: a suavização da série temporal (que ajusta o nível da série) e a suavização da tendência (que ajusta a inclinação da série ao longo do tempo). Esse método é especialmente útil para dados que apresentam padrões de crescimento ou declínio ao longo do tempo, sendo amplamente aplicado em áreas como previsão de vendas, demanda e, neste contexto, emissões de gases de efeito estufa.

A metodologia aplicada no modelo Holt envolveu a preparação dos dados, onde a coluna "ds" foi convertida para o formato de data e utilizada como índice do DataFrame, facilitando a análise temporal das emissões. O modelo de suavização exponencial de Holt foi então ajustado aos dados de emissões, utilizando uma tendência aditiva, sem considerar sazonalidade e com a tendência não amortecida. Após o ajuste do modelo, as previsões para os próximos 10 anos foram geradas, permitindo a projeção das emissões para o futuro.

Figura 5 - Previsões de Emissões com o Modelo Holt



5. Resultados e discussão

Os modelos foram aplicados para prever as emissões de gases de efeito estufa (GEE) nos próximos 10 anos, com foco na variável "emissao_normalizada". A tabela a seguir resume as métricas de desempenho obtidas por cada modelo:

Tabela 2 - Métricas de Desempenho por Modelo

MODELO	MAE	RMSE	MAPE	R²
ARIMA	150.998	592937	Não Disponível	-0,04
Prophet	13.567	4.486	0,52%	Não avaliado
Suavização de Holt	5.694	8.091	0,2%	Não avaliado

Fonte: Os autores (2024)

ARIMA

O modelo ARIMA apresentou o desempenho mais limitado entre os três métodos testados. Apesar de confirmar a estacionariedade da série temporal por meio do teste ADF (estatística: -30.166; p-valor: 0.0), as métricas de desempenho indicaram baixa aderência aos dados históricos, com R2R2 negativo (-0.04). As previsões realizadas para os próximos 10 anos apontaram um leve crescimento nas emissões até 2026, seguido por uma estabilização e posterior queda.

Tabela 3 - Desempenho Modelo ARIMA

Ano	Previsão de Emissão (toneladas)
2020	237.977
2021	247.503
2022	255.540
2023	262.976
2024	267.356
2025	269.509
2026	269.569
2027	268.514
2028	266.946
2029	265.454

Fonte: Os Autores (2024)

Esses resultados sugerem que o ARIMA pode ter limitações em capturar as complexidades da série, especialmente flutuações mais sutis ou comportamentos não lineares.

Prophet

O modelo Prophet apresentou métricas significativamente melhores, com MAE de 13.567 e RMSE de 4.486, destacando-se como uma alternativa mais precisa ao ARIMA. As previsões indicaram uma tendência de queda contínua nas emissões ao longo do período projetado, refletindo possivelmente o impacto de políticas ambientais mais rigorosas ou mudanças estruturais em setores econômicos emissores.

Suavização de Holt

A Suavização de Holt foi o modelo com melhor desempenho, com as menores métricas de erro: MAE de 5.694, RMSE de 8.091 e MAPE de 0.200. Esse modelo

projetou um cenário oposto ao do Prophet, indicando um aumento gradual nas emissões ao longo do período analisado. Esse comportamento pode estar relacionado à extrapolação de tendências recentes, sugerindo uma possível retomada de atividades econômicas ou redução na efetividade de políticas de mitigação.

O fato de os melhores resultados serem superiores sugere que o modelo de Holt capturou a maioria das tendências e variações na série temporal muito bem. A eficácia aqui pode ser devido à maneira como este modelo pode ajustar independentemente a suavização para o nível e a tendência; isso pode ser o que o torna adequado quando a série tem apenas mudanças graduais ou um padrão consistente ao longo do tempo.

Ao contrário dos outros modelos, o Holt previu um aumento no nível de emissões durante os próximos anos, fato que pode ser compreendido em relação à ênfase que ele coloca na continuação da tendência manifestada nos últimos anos pela série histórica. Esta projeção considera que, sob a continuidade dos padrões de emissões passados, fatores como o aumento das atividades econômicas emissoras ou a desaceleração na implementação de políticas de mitigação podem levar à retomada do crescimento. Esta previsão ressalta a importância de intervenções mais efetivas para mudar esse comportamento e evitar uma nova escalada nas emissões de gases de efeito estufa.

Os resultados obtidos com os modelos testados apontam para diferenças significativas na capacidade de cada abordagem de capturar tendências nas séries temporais de emissões de GEE. O ARIMA teve pior desempenho, com métricas como MAE e RMSE significativamente altas e um R^2 negativo, denotando baixa aderência aos dados históricos. As previsões para os próximos 10 anos indicaram que houve um ligeiro aumento nas emissões até 2026, após o qual se estabilizariam e então declinariam — uma possível deficiência do modelo em capturar as flutuações menores ou padrões não lineares nas séries temporais.

Por outro lado, Prophet e a suavização de Holt tiveram desempenho com maior precisão. Entre elas, a Prophet foi uma série muito peculiar porque previu uma tendência descendente das emissões, provavelmente refletindo avanços em políticas ambientais ou mudanças estruturais nos setores emissores. No entanto, as melhores métricas de desempenho, com uma melhor captura dos padrões e um aumento lento nas emissões, provavelmente devido à extrapolação de tendências recentes na recuperação econômica, foram fornecidas pelo modelo de Holt. Juntos, esses modelos fornecem cenários complementares que trazem uma visão abrangente e estratégica das diferentes trajetórias de emissões possíveis nos próximos anos.

5. Conclusão

Os resultados obtidos durante este projeto lançam luz sobre a relevância e os desafios de abordar uma questão tão crítica quanto as emissões de GEE. Este trabalho permitiu identificar a tendência histórica e projetar cenários futuros, e também mostrou que os setores de Mudança de Uso da Terra e Florestas são de onde as maiores emissões brasileiras serão originadas, permanecendo, portanto, alinhados com as premissas dos objetivos iniciais. As análises realizadas, com base em modelos como ARIMA, Prophet e a suavização de Holt, fornecem visões importantes para apoiar políticas públicas e esforços estratégicos para redução de emissões.

No entanto, algumas limitações nas análises devem ser consideradas. Modelos como ARIMA não capturaram a complexidade dos dados. Essa perda de precisão pode ser porque as séries temporais são intrinsecamente variáveis ou talvez porque alguns fatores exógenos não foram incluídos no modelo. Prophet e Holt mostraram mais fidelidade, mas ainda faltam elementos que considerem dinâmicas socioeconômicas ou ambientais mais complexas. Todavia, os resultados obtidos fornecem uma plataforma real para embasar o estabelecimento de políticas públicas que convergem com os ODS, especialmente o ODS 13 referente à ação para combater as mudanças climáticas.

É bastante evidente a partir do objetivo do projeto que ele apontará identicamente os maiores emissores setoriais, analisará as tendências preditivas e prosseguirá propondo medidas estratégicas. Porém, algumas melhorias adicionais são possíveis usando dados mais recentes e considerando variáveis complementares, como fatores climáticos e econômicos, empregando modelos ainda mais robustos, como redes neurais ou abordagens híbridas. Este trabalho não apenas contribui para a causa mundial de combate às mudanças climáticas, mas, mais importante, reitera a necessidade de colaboração entre governos, empresas e sociedade para um futuro sustentável, causando, portanto, um impacto crítico e duradouro para combater o aquecimento global.

6. Referências bibliográficas

AGÊNCIA BRASIL. Brasil registra alta na emissão de gases de efeito estufa. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2023-03/brasil-registra-alta-na-emissao-de-gases-de-efeito-estufa>. Acesso em: 3 set. 2024.

AGÊNCIA BRASIL. Mudanças climáticas propiciam a expansão de doenças como dengue, diz WWF. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2018-05/mudancas-climaticas-propiciam-expansao-de-doencas-como-dengue-diz-wwf>. Acesso em: 3 set. 2024.

BISOGNIN, Cleber; WERNER, Liane. Emissão de dióxido de carbono: um estudo em três países utilizando combinação de previsões. In: POISSON, E. Gestão da produção em foco: volume 9. 2018. Disponível em: <https://doi.org/10.5935/978-85-93729-51-5.2018B001>. Acesso em: 20 set. 2024. Capítulo 2, p. 16.

ENERGIA E AMBIENTE. COP 28: Brasil emitiu 2,3 bilhões de toneladas brutas de gases de efeito estufa em 2022. Disponível em: <https://energiaeambiente.org.br/cop-28-brasil-emitiu-23-bilhoes-de-toneladas-brutas-de-gases-de-efeito-estufa-em-2022-20231209>. Acesso em: 3 set. 2024.

ESTIMATIVAS ANUAIS DE EMISSÕES DE GASES DE EFEITO ESTUFA. [s.l: s.n.]. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/sirene/publicacoes/estimativas-anuais-de-emissoes-gee/arquivos/6a-ed-estimativas-anuais.pdf>. Acesso em: 4 set. 2024.

INSTITUTO DE PESQUISA AMBIENTAL (IPAM). Quais serão os impactos prováveis destas mudanças no nível global? Disponível em: <https://ipam.org.br/entenda/quais-serao-os-impactos-provaveis-destas-mudancas-no-nivel-global/>. Acesso em: 4 set. 2024.

IPCC. INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. Climate Change 2021: The Physical Science Basis. Cambridge University Press, 2021. Disponível em: <https://www.ipcc.ch/report/ar6/wg1/>. Acesso em: 4 set. 2024.

LUO, Y. CO2 Emission Prediction Based on Prophet, ARIMA and LSTM. Highlights in Science, Engineering and Technology, v. 76, p. 385-390, 2023. DOI: <10.54097/4k6yfr37>. Disponível em: https://www.researchgate.net/publication/377743350_CO2_Emission_Prediction_Based_on_Prophet_ARIMA_and_LSTM. Acesso em: 20 set. 2024.

MATTEI, Taise Fátima; CUNHA, Marina Silva da. Emissão de gases do efeito estufa no Brasil: análise de curto e longo prazo. Revista Agronegócio e Meio Ambiente, Maringá, PR, 2021.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de séries temporais: modelos lineares univariados. 2018. 1. ed. São Paulo: Blucher, 2018. ISBN 9788521213529. Disponível em: <https://books.google.com.br/books?id=UwC5DwAAQBAJ>. Acesso em: 25 out.2024.

POTENZA, R. F. et al. Análise das emissões brasileiras de gases de efeito estufa e suas implicações para as metas climáticas do Brasil 1970-2020. Brasília, DF: SEEG, 2021. Disponível em: <https://energiaambiente.org.br/wp-content/uploads/2023/04/SEEG-10-anos-v5.pdf>. Acesso em: 3 set. 2024.

POTENZA, Renata Fragoso; QUINTANA, Gabriel de Oliveira; CARDOSO, Anderson Matheus et al. Análise das emissões brasileiras e suas implicações para as metas climáticas do Brasil: 1970 – 2020. 2021. Disponível em: https://www.oc.eco.br/wp-content/uploads/2021/10/OC_03_relatorio_2021_FINAL.pdf#:~:text=efeito%20estufa%20no%20Brasil%20em%202020.%20No,grande%20emissor%20do%20planeta%20a%20verificar%20alta. Acesso em: 26 out.2024.

SEEG. SEEG: 10 anos de monitoramento das emissões de gases de efeito estufa. 2023. Disponível em: <https://www.oc.eco.br/wp-content/uploads/2023/03/SEEG-10-anos-v4.pdf>. Acesso em: 2 set. 2024.

UNEP. Você sabe como os gases de efeito estufa aquecem o planeta? 2023. Disponível em: <https://www.unep.org/pt-br/noticias-e-reportagens/reportagem/voce-sabe-como-os-gases-de-efeito-estufa-aquecem-o-planeta>. Acesso em: 3 set. 2024.

UNITED NATIONS ENVIRONMENT PROGRAMME. Emissions Gap Report 2023: Broken Record – Temperatures hit new highs, yet world fails to cut emissions (again).